

EXHIBIT 2

Behavioral Treatment and Normal Educational and Intellectual Functioning in Young Autistic Children

O. Ivar Lovaas
University of California, Los Angeles

Autism is a serious psychological disorder with onset in early childhood. Autistic children show minimal emotional attachment, absent or abnormal speech, retarded IQ, ritualistic behaviors, aggression, and self-injury. The prognosis is very poor, and medical therapies have not proven effective. This article reports the results of behavior modification treatment for two groups of similarly constituted, young autistic children. Follow-up data from an intensive, long-term experimental treatment group ($n = 19$) showed that 47% achieved normal intellectual and educational functioning, with normal-range IQ scores and successful first grade performance in public schools. Another 40% were mildly retarded and assigned to special classes for the language delayed, and only 10% were profoundly retarded and assigned to classes for the autistic/retarded. In contrast, only 2% of the control-group children ($n = 40$) achieved normal educational and intellectual functioning; 45% were mildly retarded and placed in language-delayed classes, and 53% were severely retarded and placed in autistic/retarded classes.

Kanner (1943) defined autistic children as children who exhibit (a) serious failure to develop relationships with other people before 30 months of age, (b) problems in development of normal language, (c) ritualistic and obsessional behaviors ("insistence on sameness"), and (d) potential for normal intelligence. A more complete behavioral definition has been provided elsewhere (Lovaas, Koegel, Simmons, & Long, 1973). The etiology of autism is not known, and the outcome is very poor. In a follow-up study on young autistic children, Rutter (1970) reported that only 1.5% of his group ($n = 63$) had achieved normal functioning. About 35% showed fair or good adjustment, usually required some degree of supervision, experienced some difficulties with people, had no personal friends, and showed minor oddities of behavior. The majority (more than 60%) remained severely handicapped and were living in hospitals for mentally retarded or psychotic individuals or in other protective settings. Initial IQ scores appeared stable over time. Other studies (Brown, 1969; DeMyer et al., 1973; Eisenberg, 1956; Freeman, Ritvo, Needleman, & Yokota, 1985; Havelkova, 1968) re-

port similar data. Higher scores on IQ tests, communicative speech, and appropriate play are considered to be prognostic of better outcome (Lotter, 1967).

Medically and psychodynamically oriented therapies have not proven effective in altering outcome (DeMyer, Hingtgen, & Jackson, 1981). No abnormal environmental etiology has been identified within the children's families (Lotter, 1967). At present, the most promising treatment for autistic persons is behavior modification as derived from modern learning theory (DeMyer et al., 1981). Empirical results from behavioral intervention with autistic children have been both positive and negative. On the positive side, behavioral treatment can build complex behaviors, such as language, and can help to suppress pathological behaviors, such as aggression and self-stimulatory behavior. Clients vary widely in the amount of gains obtained but show treatment gains in proportion to the time devoted to treatment. On the negative side, treatment gains have been specific to the particular environment in which the client was treated, substantial relapse has been observed at follow-up, and no client has been reported as recovered (Lovaas et al., 1973).

The present article reports a behavioral-intervention project (begun in 1970) that sought to maximize behavioral treatment gains by treating autistic children during most of their waking hours for many years. Treatment included all significant persons in all significant environments. Furthermore, the project focused on very young autistic children (below the age of 4 years) because it was assumed that younger children would be less likely to discriminate between environments and therefore more likely to generalize and to maintain their treatment gains. Finally, it was assumed that it would be easier to successfully mainstream a very young autistic child into preschool than it would be to mainstream an older autistic child into primary school.

It may be helpful to hypothesize an outcome of the present study from a developmental or learning point of view. One may assume that normal children learn from their everyday environ-

This study was supported by Grant MH-11440 from the National Institute of Mental Health. Aspects of this study were presented at the 1982 convention of the American Psychological Association, Washington, DC, by Andrea Ackerman, Paula Firestone, Gayle Goldstein, Ronald Leaf, John McEachin, and the author. The author expresses his deep appreciation to the many undergraduate students at the University of California, Los Angeles, who served as student therapists on the project, to the many graduate students who served as clinic supervisors, and to the many parents who trusted their children to our care. Special thanks to Laura Schreibman and Robert Koegel, who collaborated in the early stages of this research project. Donald Baer, Bruce Baker, Bradley Bucher, Arthur Woodward, and Haikang Shen provided statistical advice and help in manuscript preparation. B. J. Freeman's help in arranging access to Control Group 2 data is also appreciated.

Correspondence concerning this article should be addressed to O. Ivar Lovaas, Psychology Department, University of California, 405 Hilgard Avenue, Los Angeles, California 90024.

O. IVAR LOVAAS

ments most of their waking hours. Autistic children, conversely, do not learn from similar environments. We hypothesized that construction of a special, intense, and comprehensive learning environment for very young autistic children would allow some of them to catch up with their normal peers by first grade.

Method

Subjects

Subjects were enrolled for treatment if they met three criteria: (a) independent diagnosis of autism from a medical doctor or a licensed PhD psychologist, (b) chronological age (CA) less than 40 months if mute and less than 46 months if echolalic, and (c) prorated mental age (PMA) of 11 months or more at a CA of 30 months. The last criterion excluded 15% of the referrals.

The clinical diagnosis of autism emphasized emotional detachment, extreme interpersonal isolation, little if any toy or peer play, language disturbance (mutism or echolalia), excessive rituals, and onset in infancy. The diagnosis was based on a structured psychiatric interview with parents, on observations of the child's free-play behaviors, on psychological testing of intelligence, and on access to pediatric examinations. Over the 15 years of the project, the exact wording of the diagnosis changed slightly in compliance with changes in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III; American Psychiatric Association, 1980). During the last years, the diagnosis was made in compliance with DSM-III criteria (p. 87). In almost all cases, the diagnosis of autism had been made prior to family contact with the project. Except for one case each in the experimental group and Control Group 1, all cases were diagnosed by staff of the Department of Child Psychiatry, University of California, Los Angeles (UCLA) School of Medicine. Members of that staff have contributed to the writing of the DSM-III and to the diagnosis of autism adopted by the National Society for Children and Adults with Autism. If the diagnosis of autism was not made, the case was referred elsewhere. In other words, the project did not select its cases. More than 90% of the subjects received two or more independent diagnoses, and agreement on the diagnosis of autism was 100%. Similarly high agreement was not reached for subjects who scored within the profoundly retarded range on intellectual functioning (PMA < 11 months); these subjects were excluded from the study.

Treatment Conditions

Subjects were assigned to one of two groups: an intensive-treatment experimental group ($n = 19$) that received more than 40 hours of one-to-one treatment per week, or the minimal-treatment Control Group 1 ($n = 19$) that received 10 hours or less of one-to-one treatment per week. Control Group 1 was used to gain further information about the rate of spontaneous improvement in very young autistic children, especially those selected by the same agency that provided the diagnostic work-up for the intensive-treatment experimental group. Both treatment groups received treatment for 2 or more years. Strict random assignment (e.g., based on a coin flip) to these groups could not be used due to parent protest and ethical considerations. Instead, subjects were assigned to the experimental group unless there was an insufficient number of staff members available to render treatment (an assessment made prior to contact with the family). Two subjects were assigned to Control Group 1 because they lived further away from UCLA than a 1-hr drive, which made sufficient staffing unavailable to those clients. Because fluctuations in staff availability were not associated in any way with client characteristics, it was assumed that this assignment would produce unbiased groups. A large number of pretreatment measures were collected to test this assumption. Subjects did not change group assignment. Except for two families who left the experimental group within the first 6 months

(this group began with 21 subjects), all families stayed with their groups from beginning to end.

Assessments

Pretreatment mental age (MA) scores were based on the following scales (in order of the frequency of their use): the Bayley Scales of Infant Development (Bayley, 1955), the Cattell Infant Intelligence Scale (Cattell, 1960), the Stanford-Binet Intelligence Scale (Thorndike, 1972), and the Gesell Infant Development Scale (Gesell, 1949). The first three scales were administered to 90% of the subjects, and relative usage of these scales was similar in each group. Testing was carried out by graduate students in psychology who worked under the supervision of clinical psychologists at UCLA or licensed PhD psychologists at other agencies. The examiner chose the test that would best accommodate each subject's developmental level, and this decision was reached independently of the project staff. Five subjects were judged to be untestable (3 in the experimental group and 2 in Control Group 1). Instead, the Vineland Social Maturity Scale (Doll, 1953) was used to estimate their MAs (with the mother as informant). To adjust for variations in MA scores as a function of the subject's CA at the time of test administration, PMA scores were calculated for a CA at 30 months ($MA/CA \times 30$).

Behavioral observations were based on videotaped recordings of the subject's free-play behavior in a playroom equipped with several simple early-childhood toys. These videotaped recordings were subsequently scored for amount of (a) *self-stimulatory behaviors*, defined as prolonged ritualistic, repetitive, and stereotyped behavior such as body-rocking, prolonged gazing at lights, excessive hand-flapping, twirling the body as a top, spinning or lining of objects, and licking or smelling of objects or wall surfaces; (b) *appropriate play behaviors*, defined as those limiting the use of toys in the playroom to their intended purposes, such as pushing the truck on the floor, pushing buttons on the toy cash register, putting a record on the record player, and banging with the toy hammer; and (c) *recognizable words*, defined to include any recognizable word, independent of whether the subject used it in a meaningful context or for communicative purposes. One observer who was naive about subjects' group placement scored all tapes after being trained to agree with two experienced observers (using different training tapes from similar subjects). Interobserver reliability was scored on 20% of the tapes (randomly selected) and was computed for each category of behavior for each subject by dividing the sum of observer agreements by the sum of agreements and disagreements. These scores were then summed and averaged across subjects. The mean agreement (based both on occurrences and nonoccurrences) was 91% for self-stimulatory behavior, 85% for appropriate play behavior, and 100% for recognizable words. A more detailed description of these behavioral recordings has been provided elsewhere (Lovaas et al., 1973).

A 1-hr parent interview about the subjects' earlier history provided some diagnostic and descriptive information. Subjects received a score of 1 for each of the following variables parents reported: no recognizable words; no toy play (failed to use toys for their intended function); lack of emotional attachment (failed to respond to parents' affection); apparent sensory deficit (parents had suspected their child to be blind or deaf because the child exhibited no or minimal eye contact and showed an unusually high pain threshold); no peer play (subject did not show interactive play with peers); self-stimulatory behavior; tantrums (aggression toward family members or self); and no toilet training. These 8 measures from parents' intake interviews were summed to provide a sum pathology score. The intake interview also provided information about abnormal speech (0 = normal and meaningful language, however limited; 1 = echolalic language used meaningfully [e.g., to express needs]; 2 = echolalia; and 3 = mute); age of walking; number of siblings in the family; socioeconomic status of the father; sex; and neurological examinations (including EEGs and CAT scans) that resulted in findings of pathology. Finally, CA at first diagnosis and at the beginning of the

TREATMENT OF AUTISTIC CHILDREN

5

present treatment were recorded. This yielded a total of 20 pretreatment measures, 8 of which were collapsed into 1 measure (sum pathology).

A brief clinical description of the experimental group at intake follows (identical to that for Control Group 1): Only 2 of the 19 subjects obtained scores within the normal range of intellectual functioning; 7 scored in the moderately retarded range, and 10 scored in the severely retarded range. No subject evidenced pretend or imaginary play, only 2 evidenced *complex* (several different or heterogeneous behaviors that together formed one activity) play, and the remaining subjects showed *simple* (the same elementary but appropriate response made repeatedly) play. One subject showed minimal appropriate speech, 7 were echolalic, and 11 were mute. According to the literature that describes the developmental delays of autistic children in general, the autistic subjects in the present study constituted an average (or below average) sample of such children.

Posttreatment measures were recorded as follows: Between the ages of 6 and 7 years (when a subject would ordinarily have completed first grade), information about the subjects' first-grade placement was sought and validated; about the same time, an IQ score was obtained. Testing was carried out by examiners who were naive about the subjects' group placement. Different scales were administered to accommodate different developmental levels. For example, a subject with a regular educational placement received a Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) or a Stanford-Binet Intelligence Scale (Thorndike, 1972), whereas a subject in an autistic/retarded class received a nonverbal test like the Merrill-Palmer Pre-School Performance Test (Stutsman, 1948). In all instances of subjects having achieved a normal IQ score, the testing was eventually replicated by other examiners. The scales (in order of the frequency of usage) included the WISC-R (Wechsler, 1974), the Stanford-Binet (Thorndike, 1972), the Peabody Picture Vocabulary Test (Dunn, 1981), the Wechsler Pre-School Scale (Wechsler, 1967), the Bayley Scales of Infant Development (Bayley, 1955), the Cattell Infant Intelligence Scale (Cattell, 1960), and the Leiter International Performance Scale (Leiter, 1959). Subjects received a score of 3 for *normal functioning* if they received a score on the WISC-R or Stanford-Binet in the normal range, completed first grade in a normal class in a school for normal children, and were advanced to the second grade by the teacher. Subjects received a score of 2 if they were placed in first-grade in a smaller *aphasia* (language delayed, language handicapped, or learning disabled) class. Placement in the *aphasia* class implied a higher level of functioning than placement in classes for the autistic/retarded, but the diagnosis of autism was almost always retained. A score of 1 was given if the first-grade placement was in a class for the autistic/retarded and if the child's IQ score fell within the severely retarded range.

Treatment Procedure

Each subject in the experimental group was assigned several well-trained student therapists who worked (part-time) with the subject in the subject's home, school, and community for an average of 40 hr per week for 2 or more years. The parents worked as part of the treatment team throughout the intervention; they were extensively trained in the treatment procedures so that treatment could take place for almost all of the subjects' waking hours, 365 days a year. A detailed presentation of the treatment procedure has been presented in a teaching manual (Lovasz et al., 1980). The conceptual basis of the treatment was reinforcement (operant) theory; treatment relied heavily on discrimination-learning data and methods. Various behavioral deficiencies were targeted, and separate programs were designed to accelerate development for each behavior. High rates of aggressive and self-stimulatory behaviors were reduced by being ignored; by the use of time-out; by the shaping of alternate, more socially acceptable forms of behavior; and (as a last resort) by the delivery of a loud "no" or a slap on the thigh contingent upon the presence of the undesirable behavior. Contingent physical aversives were not used in the control group because inadequate staffing

in that group did not allow for adequate teaching of alternate, socially appropriate behaviors.

During the first year, treatment goals consisted of reducing self-stimulatory and aggressive behaviors, building compliance to elementary verbal requests, teaching imitation, establishing the beginnings of appropriate toy play, and promoting the extension of the treatment into the family. The second year of treatment emphasized teaching expressive and early abstract language and interactive play with peers. Treatment was also extended into the community to teach children to function within a preschool group. The third year emphasized the teaching of appropriate and varied expression of emotions; preacademic tasks like reading, writing, and arithmetic; and *observational learning* (learning by observing other children learn). Subjects were enrolled only in those preschools where the teacher helped to carry out the treatment program. Considerable effort was exercised to mainstream subjects in a normal (average and public) preschool placement and to avoid initial placement in special education classes with the detrimental effects of exposure to other autistic children. This occasionally entailed withholding the subject's diagnosis of autism. If the child became known as autistic (or as "a very difficult child") during the first year in preschool, the child was encouraged to enroll in another, unfamiliar school (to start fresh). After preschool, placement in public education classes was determined by school personnel. All children who successfully completed normal kindergarten successfully completed first grade and subsequent normal grades. Children who were observed to be experiencing educational and psychological problems received their school placement through Individualized Educational Plan (IEP) staffings (attended by educators and psychologists) in accordance with the Education For All Handicapped Children Act of 1975.

All subjects who went on to a normal first grade were reduced in treatment from the 40 hr per week characteristic of the first 2 years to 10 hr or less per week during kindergarten. After a subject had started first grade, the project maintained a minimal (at most) consultant relationship with some families. In two cases, this consultation and the subsequent correction of problem behaviors were judged to be essential in maintaining treatment gains. Subjects who did not recover in the experimental group received 40 hr or more per week of one-to-one treatment for more than 6 years (more than 14,000 hr of one-to-one treatment), with some improvement shown each year but with only 1 subject recovering.

Subjects in Control Group 1 received the same kind of treatment as those in the experimental group but with less intensity (less than 10 hr of one-to-one treatment per week) and without systematic physical aversives. In addition, these subjects received a variety of treatments from other sources in the community such as those provided by small special education classes.

Control Group 2 consisted of 21 subjects selected from a larger group ($N = 62$) of young autistic children studied by Freeman et al. (1985). These subjects came from the same agency that diagnosed 95% of our other subjects. Data from Control Group 2 helped to guard against the possibility that subjects who had been referred to us for treatment constituted a subgroup with particularly favorable or unfavorable outcomes. To provide a group of subjects similar to those in the experimental group and Control Group 1, subjects for Control Group 2 were selected if they were 42 months old or younger when first tested, had IQ scores above 40 at intake, and had follow-up testing at 6 years of age. These criteria resulted in the selection of 21 subjects. Subjects in Control Group 2 were treated like Control Group 1 subjects but were not treated by the Young Autism Project described here.

Results

Pretreatment Comparisons

Eight pretreatment variables from the experimental group and Control Group 1 (CA at first diagnosis, CA at onset of treat-

Table 1*Means and F Ratios From Comparisons Between Groups on Intake Variables*

Group	Diagnosis CA	Treatment CA	PMA	Recognizable words	Toy play	Self-stimulation	Sum pathology	Abnormal speech
Experimental	32.0	34.6	18.8	.42	28.2	12.1	6.9	2.4
Control 1	35.3	40.9	17.1	.58	20.2	19.6	6.4	2.2
F*	1.58	4.02*	1.49	.92	2.76	3.37	.82	.36

Note. CA = chronological age; PMA = prorated mental age. Experimental group, $n = 19$; Control Group 1, $n = 19$.

* $df = 1, 36$.

* $p < .05$.

ment, PMA, sum pathology, abnormal speech, self-stimulatory behavior, appropriate toy play, and recognizable words) were subjected to a multivariate analysis of variance (MANOVA; Brecht & Woodward, 1984). The means and F ratios from this analysis are presented in Table 1. As can be seen, there were no significant differences between the groups except for CA at onset of our treatment ($p < .05$). Control subjects were 6 months older on the average than experimental subjects (mean CAs of 35 months vs. 32 months, respectively). These differences probably reflect the delay of control subjects in their initiation into the treatment project because of staff shortages; analysis will show that differential CAs are not significantly related to outcome. To ascertain whether another test would reveal a statistically significant difference between the groups on toy play, descriptions of the subjects' toy play (taken from the videotaped recordings) were typed on cards and rated for their developmental level by psychology students who were naive about the purpose of the ratings and subject group assignment. The ratings were reliable among students ($r = .79, p < .001$), and an F test showed no significant difference in developmental levels of toy play between the two groups.

The respective means from the experimental group and Control Group 1 on the eight variables from the parent interview were .89 and .74 for sensory deficit, .63 and .42 for adult rejection, .58 and .47 for no recognizable words, .53 and .63 for no toy play, 1.0 and 1.0 for no peer play, .95 and .89 for body self-stimulation, .89 and .79 for tantrums, and .68 and .63 for no toilet training. The experimental group and Control Group 1 were also similar in onset of walking (6 vs. 8 early walkers; 1 vs. 2 late walkers), number of siblings in the family (1.26 in each group), socioeconomic status of the father (Level 49 vs. Level 54 according to 1950 Bureau of the Census standards), boys to girls (16:3 vs. 11:8); and number of subjects referred for neurological examinations (10 vs. 15) who showed signs of damage (0 vs. 1). The numbers of favorable versus unfavorable prognostic signs (directions of differences) on the pretreatment variables divide themselves equally between the groups. In short, the two groups appear to have been comparable at intake.

Follow-Up Data

Subjects' PMA at intake, follow-up educational placement, and IQ scores were subjected to a MANOVA that contrasted the experimental group with Control Groups 1 and 2. At intake, there were no significant differences between the experimental group and the control groups. At follow-up, the experimental group was significantly higher than the control groups on educa-

tional placement ($p < .001$) and IQ ($p < .01$). The two control groups did not differ significantly at intake or at follow-up. In short, data from Control Group 2 replicate those from Control Group 1 and further validate the effectiveness of our experimental treatment program. Data are given in Table 2 that show the group means from pretreatment PMA and posttreatment educational placement and IQ scores. The table also shows the F ratios and significance levels of the three group comparisons.

In descriptive terms, the 19-subject experimental group shows 9 children (47%) who successfully passed through normal first grade in a public school and obtained an average of above average score on IQ tests ($M = 107$, range = 94–120). Eight subjects (42%) passed first grade in aphasia classes and obtained a mean IQ score within the mildly retarded range of intellectual functioning ($M = 70$, range = 56–95). Only two children (10%) were placed in classes for autistic/retarded children and scored in the profoundly retarded range (IQ < 30).

There were substantial increases in the subjects' levels of intellectual functioning after treatment. The experimental group subjects gained on the average of 30 IQ points over Control Group 1 subjects. Thus the number of subjects who scored within the normal range of intellectual functioning increased from 2 to 12, whereas the number of subjects within the moderate-to-severe range of intellectual retardation dropped from 10 to 3. As of 1986, the achievements of experimental group sub-

Table 2*Means and F Ratios for Measures at Pretreatment and Posttreatment*

Group	Intake PMA	Follow-up	
		EDP	IQ
Means			
Experimental	18.8	2.37	83.3
Control 1	17.1	1.42	52.2
Control 2	17.6	1.57	57.5
<i>F</i> ratios*			
Experimental × Control 1	1.47	23.6**	14.4**
Experimental × Control 2	0.77	17.6**	10.4**
Control 1 × Control 2	0.14	0.63	0.45

Note. PMA = prorated mental age; EDP = educational placement. Experimental group, $n = 19$; Control Group 1, $n = 19$; Control Group 2, $n = 21$.

* $df = 1, 56$.

* $p < .01$. ** $p < .001$.

TREATMENT OF AUTISTIC CHILDREN

Table 3
*Educational Placement and Mean
 and Range of IQ at Follow-Up*

Group	Recovered	Aphasic	Autistic/Retarded
Experimental			
<i>N</i>	9	8	2
<i>MIQ</i>	107	70	30
Range	94-120	56-95	—*
Control Group 1			
<i>N</i>	0	8	11
<i>MIQ</i>	—	74	36
Range	—	30-102	20-73
Control Group 2			
<i>N</i>	1	10	10
<i>MIQ</i>	99	67	44
Range	—	49-81	35-54

Note. Dashes indicate no score or no entry.

* Both children received the same score.

jects have remained stable. Only 2 subjects have been reclassified: 1 subject (now 18 years old) was moved from an aphasia to a normal classroom after the sixth grade; 1 subject (now 13 years old) was moved from an aphasia to an autistic/retarded class placement.

The MA and IQ scores of the two control groups remained virtually unchanged between intake and follow-up, consistent with findings from other studies (Freeman et al., 1985; Rutter, 1970). The stability of the IQ scores of the young autistic children, as reported in the Freeman et al. study, is particularly relevant for the present study because it reduces the possibility of spontaneous recovery effects. In descriptive terms, the combined follow-up data from the control groups show that their subjects fared poorly: Only 1 subject (2%) achieved normal functioning as evidenced by normal first-grade placement and an IQ of 99 on the WISC-R; 18 subjects (45%) were in aphasia classes (mean IQ = 70, range = 30-101); and 21 subjects (53%) were in classes for the autistic/retarded (mean IQ = 40, range = 20-73). Table 3 provides a convenient descriptive summary of the main follow-up data from the three groups.

One final control procedure subjected 4 subjects in the experimental group (Ackerman, 1980) and 4 subjects in Control Group 1 (McEachin & Leaf, 1984) to a treatment intervention in which one component of treatment (the loud "no" and occasional slap on the thigh contingent on self-stimulatory, aggressive, and noncompliant behavior) was at first withheld and then introduced experimentally. A within-subjects replication design was used across subjects, situations, and behaviors, with baseline observations varying from 3 weeks to 2 years after treatment had started (using contingent positive reinforcement only). During baseline, when the contingentaversive component was absent, small and unstable reductions were observed in the large amount of inappropriate behaviors, and similar small and unstable increases were observed in appropriate behaviors such as play and language. These changes were insufficient to allow for the subjects' successful mainstreaming. Introduction of contingent aversives resulted in a sudden and stable reduction in the inappropriate behaviors and a sudden and stable increase in appropriate behaviors. This experimental intervention helps to establish two points: First, at least one compo-

nent in the treatment program functioned to produce change, which helps to reduce the effect of placebo variables. Second, this treatment component affected both the experimental and control groups in a similar manner, supporting the assumption that the two groups contained similar subjects.

Analyses of variance were carried out on the eight pretreatment variables to determine which variables, if any, were significantly related to outcome (gauged by educational placement and IQ) in the experimental group and Control Group 1. Pro-rated mental age was significantly ($p < .03$) related to outcome in both groups, a finding that is consistent with reports from other investigators (DeMyer et al., 1981). In addition, abnormal speech was significantly ($p < .01$) related to outcome in Control Group 1. Chronological age at onset of our treatment was not related to outcome, which is important because the two groups differed significantly on this variable at intake (by 6 months). The failure of CA to relate to outcome may be based on the very young age of all subjects at onset of treatment.

Conceivably, a linear combination of pretreatment variables could have predicted outcome in the experimental group. Using a discriminant analysis (Ray, 1982) with the eight variables used in the first multivariate analysis, it was possible to predict perfectly the 9 subjects who did achieve normal functioning, and no subject was predicted to achieve this outcome who did not. In this analysis, PMA was the only variable that was significantly related to outcome. Finally, when this prediction equation was applied to Control Group 1 subjects, 8 were predicted to achieve normal functioning with intensive treatment: this further verifies the similarity between the experimental group and Control Group 1 prior to treatment.

Discussion

This article reports the results of intensive behavioral treatment for young autistic children. Pretreatment measures revealed no significant differences between the intensively treated experimental group and the minimally treated control groups. At follow-up, experimental group subjects did significantly better than control group subjects. For example, 47% of the experimental group achieved normal intellectual and educational functioning in contrast to only 2% of the control group subjects.

The study incorporated certain methodological features designed to increase confidence in the effectiveness of the experimental group treatment:

- Pretreatment differences between the experimental and control groups were minimized in four ways. First, the assignment of subjects to groups was as random as was ethically possible. The assignment apparently produced unbiased groups as evidenced by similar scores on the 20 pretreatment measures and by the prediction that an equal number of Control Group 1 and experimental group subjects would have achieved normal functioning had the former subjects received intensive treatment. Second, the experimental group was not biased by receiving subjects with a favorable diagnosis or biased IQ testing because both diagnosis and IQ tests were constant across groups. Third, the referral process did not favor the project cases because there were no significant differences between Control Groups 1 and 2 at intake or follow-up, even though Control Group 2 subjects were referred to others by the same agency.

Fourth, subjects stayed within their groups, which preserved the original (unbiased) group assignment.

2. A favorable outcome could have been caused not by the experimental treatment but by the attitudes and expectations of the staff. There are two findings that contradict this possibility of treatment agency (placebo) effects. First, because Control Group 2 subjects had no contact with the project, and because there was no difference between Control Groups 1 and 2 at follow-up, placebo effects appear implausible. Second, the within-subjects study showed that at least one treatment component contributed to the favorable outcome in the intensive treatment (experimental) group.

3. It may be argued that the treatment worked because the subjects were not truly autistic. This is counterindicated by the high reliability of the independent diagnosis and by the outcome data from the control groups, which are consistent with those reported by other investigators (Brown, 1969; DeMeyer et al., 1973; Eisenberg, 1956; Freeman et al., 1985; Havelkova, 1968; Rutter, 1970) for groups of young autistic children diagnosed by a variety of other agencies.

4. The spontaneous recovery rate among very young autistic children is unknown, and without a control group the favorable outcome in the experimental group could have been attributed to spontaneous recovery. However, the poor outcome in the similarly constituted Control Groups 1 and 2 would seem to eliminate spontaneous recovery as a contributing factor to the favorable outcome in the experimental group. The stability of the IQ test scores in the young autistic children examined by Freeman et al. (1985) attests once again to the chronicity of autistic behaviors and serves to further negate the effects of spontaneous recovery.

5. Posttreatment data showed that the effects of treatment (a) were substantial and easily detected, (b) were apparent on comprehensive, objective, and socially meaningful variables (IQ and school placement), and (c) were consistent with a very large body of prior research on the application of learning theory to the treatment and education of developmentally disabled persons and with the very extensive (100-year-old) history of psychology laboratory work on learning processes in man and animals. In short, the favorable outcome reported for the intensive-treatment experimental group can in all likelihood be attributed to treatment.

A number of measurement problems remain to be solved. For example, play, communicative speech, and IQ scores define the characteristics of autistic children and are considered predictors of outcome. Yet the measurement of these variables is no easy task. Consider play. First, play undoubtedly varies with the kinds of toys provided. Second, it is difficult to distinguish low levels of toy play (simple and repetitive play associated with young, normal children) from high levels of self-stimulatory behavior (a psychotic attribute associated with autistic children). Such problems introduce variability that needs immediate attention before research can proceed in a meaningful manner.

The term *normal functioning* has been used to describe children who successfully passed normal first grade and achieved an average IQ on the WISC-R. But questions can be asked about whether these children truly recovered from autism. On the one hand, educational placement is a particularly valuable measure of progress because it is sensitive to both educational accomplishments and social-emotional functions. Also, continual

promotion from grade to grade is made not by one particular teacher but by several teachers. School personnel describe these children as indistinguishable from their normal friends. On the other hand, certain residual deficits may remain in the normal functioning group that cannot be detected by teachers and parents and can only be isolated on closer psychological assessment, particularly as these children grow older. Answers to such questions will soon be forthcoming in a more comprehensive follow-up (McEachin, 1987).

Several questions about treatment remain. It is unlikely that a therapist or investigator could replicate our treatment program for the experimental group without prior extensive theoretical and supervised practical experience in one-to-one behavioral treatment with developmentally disabled clients as described here and without demonstrated effectiveness in teaching complex behavioral repertoires as in imitative behavior and abstract language. In the within-subjects studies that were reported, contingent aversives were isolated as one significant variable. It is therefore unlikely that treatment effects could be replicated without this component. Many treatment variables are left unexplored, such as the effect of normal peers. Furthermore, the successful mainstreaming of a 2-4-year-old into a normal preschool group is much easier than the mainstreaming of an older autistic child into the primary grades. This last point underscores the importance of early intervention and places limits on the generalization of our data to older autistic children.

Historically, psychodynamic theory has maintained a strong influence on research and treatment with autistic children, offering some hope for recovery through experiential manipulations. By the mid-1960s, an increasing number of studies reported that psychodynamic practitioners were unable to deliver on that promise (Rimland, 1964). One reaction to those failures was an emphasis on organic theories of autism that offered little or no hope for major improvements through psychological and educational interventions. In a comprehensive review of research on autism, DeMyer et al. (1981) concluded that "[in the past] psychotic children were believed to be potentially capable of normal functioning in virtually all areas of development . . . during the decade of the 1970s it was the rare investigator who even gave lip-service to such previously held notions . . . infantile autism is a type of developmental disorder accompanied by severe and, to a large extent, permanent intellectual/behavioral deficits" (p. 432).

The following points can now be made. First, at least two distinctively different groups emerged from the follow-up data in the experimental group. Perhaps this finding implies different etiologies. If so, future theories of autism will have to identify these groups of children. Second, on the basis of testing to date, the recovered children show no permanent intellectual or behavioral deficits and their language appears normal, contrary to the position that many have postulated (Rutter, 1974; Churchill, 1978) but consistent with Kanner's (1943) position that autistic children possess potentially normal or superior intelligence. Third, at intake, all subjects evidenced deficiencies across a wide range of behaviors, and during treatment they showed a broad improvement across all observed behaviors. The kind of (hypothesized) neural damage that mediates a particular kind of behavior, such as language (Rutter, 1974), is not consistent with these data.

TREATMENT OF AUTISTIC CHILDREN

9

Although serious problems remain for exactly defining autism or identifying its etiology, one encouraging conclusion can be stated: Given a group of children who show the kinds of behavioral deficits and excesses evident in our pretreatment measures, such children will continue to manifest similar severe psychological handicaps later in life unless subjected to intensive behavioral treatment that can indeed significantly alter that outcome.

These data promise a major reduction in the emotional hardships of families with autistic children. The treatment procedures described here may also prove equally effective with other childhood disorders, such as childhood schizophrenia. Certain important, practical implications in these findings may also be noted. The treatment schedule of subjects who achieved normal functioning could be reduced from 40 hr per week to infrequent visits even after the first 2 years of treatment. The assignment of one full-time special-education teacher for 2 years would cost an estimated \$40,000, in contrast to the nearly \$2 million incurred (in direct costs alone) by each client requiring life-long institutionalization.

References

- Ackerman, A. B. (1980). *The contribution of punishment to the treatment of preschool aged children*. Unpublished doctoral dissertation, University of California, Los Angeles.
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- Bayley, N. (1955). On the growth of intelligence. *American Psychologist*, 10, 805-818.
- Brecht, M. L., & Woodward, J. A. (1984). GANOVA: A univariate/multivariate analysis of variance program for the personal computer. *Educational and Psychological Measurement*, 44, 169-173.
- Brown, J. (1969). Adolescent development of children with infantile psychosis. *Seminars in Psychiatry*, 1, 79-89.
- Cattell, P. (1960). *The measurement of intelligence of infants and young children*. New York: Psychological Corporation.
- Churchill, D. W. (1978). Language: The problem beyond conditioning. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment* (pp. 71-85). New York: Plenum.
- DeMyer, M. K., Barton, S., DeMyer, W. E., Norton, J. A., Allen, J., & Steele, R. (1973). Prognosis in autism: A follow-up study. *Journal of Autism and Childhood Schizophrenia*, 3, 199-246.
- DeMyer, M. K., Hingtgen, J. N., & Jackson, R. K. (1981). Infantile autism reviewed: A decade of research. *Schizophrenia Bulletin*, 7, 388-451.
- Doll, E. A. (1953). *The measurement of social competence*. Minneapolis, MN: Minneapolis Educational Test Bureau.
- Dunn, L. M. (1981). *Peabody Picture Vocabulary Test*. Circle River, MI: American Guidance Service.
- Education for All Handicapped Children Act of 1975. Washington, DC: *Congressional Record*.
- Eisenberg, L. (1956). The autistic child in adolescence. *American Journal of Psychiatry*, 112, 607-612.
- Freeman, B. J., Ritvo, E. R., Needleman, R., & Yokota, A. (1985). The stability of cognitive and linguistic parameters in autism: A 5-year study. *Journal of the American Academy of Child Psychiatry*, 24, 290-311.
- Gesell, A. (1949). *Gesell Developmental Schedules*. New York: Psychological Corporation.
- Havelkova, M. (1968). Follow-up study of 71 children diagnosed as psychotic in preschool age. *American Journal of Orthopsychiatry*, 38, 846-857.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2, 217-250.
- Leiter, R. G. (1959). Part I of the manual for the 1948 revision of the Leiter International Performance Scale: Evidence of the reliability and validity of the Leiter tests. *Psychology Service Center Journal*, 11, 1-72.
- Lotter, V. (1967). Epidemiology of autistic conditions in young children: II. Some characteristics of the parents and children. *Social Psychiatry*, 1, 163-173.
- Lovaas, O. I., Ackerman, A. B., Alexander, D., Firestone, P., Perkins, J., & Young, D. (1980). *Teaching developmentally disabled children: The me book*. Austin, TX: Pro-Ed.
- Lovaas, O. I., Koegel, R. L., Simmons, J. Q., & Long, J. (1973). Some generalization and follow-up measures on autistic children in behavior therapy. *Journal of Applied Behavior Analysis*, 6, 131-166.
- McEachin, J. J. (1987). *Outcome of autistic children receiving intensive behavioral treatment: Residual deficits*. Unpublished doctoral dissertation, University of California, Los Angeles.
- McEachin, J. J., & Leaf, R. B. (1984, May). *The role of punishment in motivation of autistic children*. Paper presented at the convention of the Association for Behavior Analysis, Nashville, TN.
- Ray, A. A. (1982). *Statistical Analysis System user's guide: Statistics, 1982 edition*. Cary, NC: SAS Institute.
- Rimland, B. (1964). *Infantile autism*. New York: Appleton-Century-Crofts.
- Rutter, M. (1970). Autistic children: Infancy to adulthood. *Seminars in Psychiatry*, 2, 435-450.
- Rutter, M. (1974). The development of infantile autism. *Psychological Medicine*, 4, 147-163.
- Stutsman, R. (1948). *Guide for administering the Merrill-Palmer Scale of Mental Tests*. New York: Harcourt, Brace & World.
- Thorndike, R. L. (1972). *Manual for Stanford-Binet Intelligence Scale*. Boston: Houghton Mifflin.
- Wechsler, D. (1967). *Manual for the Wechsler Pre-School and Primary Scale of Intelligence*. New York: Psychological Corporation.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corporation.

Received October 10, 1985

Revision received March 28, 1986 ■

American Journal on Mental Retardation
 1993 Vol. 97 No. 4 359-372
 © 1993 American Association on Mental Retardation

Long-Term Outcome for Children With Autism Who Received Early Intensive Behavioral Treatment

John J. McEachin, Tristram Smith, and O. Ivar Lovasz
 University of California, Los Angeles

After a very intensive behavioral intervention, an experimental group of 19 preschool-age children with autism achieved less restrictive school placements and higher IQs than did a control group of 19 similar children by age 7 (Lovasz, 1987). The present study followed-up this finding by assessing subjects at a mean age of 11.5 years. Results showed that the experimental group preserved its gains over the control group. The 9 experimental subjects who had achieved the best outcomes at age 7 received particularly extensive evaluations indicating that 8 of them were indistinguishable from average children on tests of intelligence and adaptive behavior. Thus, behavioral treatment may produce long-lasting and significant gains for many young children with autism.

Infantile autism is a condition marked by severe impairment in intellectual, social, and emotional functioning. Its onset occurs in infancy, and the prognosis appears

This study was supported by Grant No. MH-11440 from the National Institute of Mental Health. The study was based on a dissertation submitted to the University of California, Los Angeles, Department of Psychology, in partial fulfillment of the requirements for the doctoral degree. The authors express their deep appreciation to the many students at UCLA who served as therapists and helped to make this study possible. Special thanks to Bruce Baker and Duane Buhrmester, who helped in the design of this study. Requests for reprints of this article, copies of the Clinical Rating Scale, or additional information about this study should be sent to O. Ivar Lovasz, 405 Hilgard Ave., UCLA, Department of Psychology, Los Angeles, CA 90024-1563.

to be extremely poor (Rutter, 1978). For example, in the longest prospective follow-up study with a sound methodological design, Rutter (1970) found that only 1 of 64 subjects with autism (fewer than 2%) could be considered free of clinically significant problems by adulthood, as evidenced by holding a job, living independently, and maintaining an active and age-appropriate social life. The remaining subjects showed numerous dysfunctions, such as marked oddities in behavior, social isolation, and florid psychopathology. The majority of subjects required supervised living conditions.

Professionals have attempted a wide variety of interventions in an effort to help children with autism. For many years, no scientific evidence showed that any of these interventions brightened the children's long-term prognosis (DeMyer et al., 1981). How-

Method

Subjects and Background

Characteristics of the subjects and their treatment have been described elsewhere (Lovaas, 1987) and will only be summarized here. The initial treatment study contained 38 children who, at the time of intake, were very young (less than 40 months if mute, less than 46 months if echolalic) and had received a diagnosis of autism from a licensed clinical psychologist or psychiatrist not involved in the study. These 38 subjects were divided into an experimental group and a control group. The assignment to groups was made on the basis of staff availability. At the beginning of each academic quarter, treatment teams were formed. The clinic director and staff members then determined whether any opening existed for intensive treatment. If so, the next referral received would enter the experimental group; otherwise, the subject entered the control group. The experimental group contained 19 children who received 40 or more hours per week of one-to-one behavioral treatment for 2 or more years. The control group was comprised of 19 children who received a much less intensive intervention (10 hours a week or less of one-to-one behavioral treatment in addition to a variety of treatments provided by community agencies, such as parent training or special education classes). The initial study also included a second control group, consisting of 21 children with autism who were followed over time by a nearby agency but who were never referred for this study. However, these 21 subjects were not available for the present investigation. On standardized measures of intelligence, the second control group did not differ from either the experimental group or the first control group at intake, nor did it differ from the first control group when evaluated again when the subjects were 7 years old. These findings suggest that, as measured by standardized tests, (a) the children with autism who were referred to us for

treatment were comparable to children with autism seen elsewhere and (b) the minimal treatment provided to the first control group did not alter intellectual functioning.

Statistical analysis of an extensive range of pretreatment measures confirmed that the experimental group and control group were comparable at intake and closely matched on such important variables as IQ and severity of disturbance. The mean chronological age (CA) at diagnosis for subjects in the experimental group was 32 months. Their mean IQ was 53 (range 30 to 82; all IQs are given as deviation scores). The mean CA of subjects in the control group was 35 months; their mean IQ was 46 (range 30 to 80). Most of the subjects were mute, all had gross deficiencies in receptive language, none played with peers or showed age-appropriate toy play, all were emotionally withdrawn, most had severe tantrums, and all showed extensive ritualistic and stereotyped (self-stimulatory) behaviors. Thus, they appeared to be a representative sample of children with autism (Lovaas, Smith, & McEachin, 1989). A more complete presentation of the intake data was reported by Lovaas (1987).

The children in the experimental group and control group received their respective treatments from trained student therapists who worked in the child's home. The parents also worked with their child, and they received extensive instruction and supervision on appropriate treatment techniques. Whenever possible, the children were integrated into regular preschools. The treatment focused primarily on developing language, increasing social behavior, and promoting cooperative play with peers along with independent and appropriate toy play. Concurrently, substantial efforts were directed at decreasing excessive rituals, tantrums, and aggressive behavior. (For a more detailed description of the intervention program, see the treatment manual [Lovaas et al., 1980] and instructional videotapes that supplement the manual [Lovaas & Leaf, 1981].)

At the time of the present follow-up (1984-1985), the mean CA of the experimen-

(1) group children was 13 years (range = 9 to 19 years). All children who had achieved normal functioning by the age of 7 years had ended treatment by that point. (Normal functioning was operationally defined as scoring within the normal range on standardized intelligence tests and successfully completing first grade in a regular, nonspecial education class entirely on one's own.) On the other hand, some of the children who had not achieved normal functioning at 7 years of age had, at the request of their parents, remained in treatment. The length of time that experimental subjects had been out of treatment ranged from 0 to 12 years (mean = 5), with the normal-functioning children having been out for 3 to 9 years (mean = 5).

The mean age of subjects in the control group was 10 years (range 6 to 14). The length of time that these children had been out of treatment ranged from 0 to 9 years (mean = 3). Thus, experimental subjects tended to be older and had been out of treatment longer than had control subjects. This difference in age occurred because the first referrals for the study were all assigned to the experimental group due to the fact that referrals came slowly (7 in the first 3.5 years) and therapists were available to treat all of them. (As noted earlier, subjects were assigned to the experimental group if therapists were available to treat them; otherwise, they entered the control group.)

Statistical analyses were conducted to test whether a bias resulted from the tendency for the first referrals to go into the experimental group. For example, it is conceivable that the first referrals could have been higher functioning at intake or could have had a better prognosis than subsequent referrals. If so, the subject assignment procedure could have favored the experimental group. To assess this possibility, we correlated the order of referral with intake IQ and with IQ at the first follow-up (age 7 years). Pearson correlations were computed across both groups and within each group. These analyses indicated that the order in which subjects were referred was not associated

with intake IQ or outcome IQ. Consequently, although the tendency for the first referrals to enter the experimental group created a potential bias, the data indicate that this was unlikely.

Procedure

The assessment procedure included ascertaining school placement and administering three standardized tests. Information on school placement was obtained from subjects' parents, who classified them as being in either a regular or a special education class (e.g., a class for children with autism or mental retardation, language delays, multihandicaps, or learning disabilities). The three standardized tests were as follows:

1. *Intelligence test.* The Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) was administered when subjects were able to provide verbal responses. This included all 9 best-outcome experimental subjects plus 8 of the remaining 10 experimental subjects and 6 of the 19 control subjects. For subjects who were not able to provide verbal responses, the Leiter International Performance Scale (Leiter, 1959) and the Peabody Picture Vocabulary Test-Revised (Dunn, 1981) were administered. All of these tests have been widely used for the assessment of intellectual functioning in children with autism (Short & Marcus, 1986).

2. *The Vineland Adaptive Behavior Scales* (Sparrow, Balla, & Cicchetti, 1984). The Vineland is a structured interview administered to parents assessing the extent to which their child exhibits behaviors that are needed to cope effectively with the everyday environment.

3. *The Personality Inventory for Children* (Wirt, Lachar, Klinedinst, & Seat, 1977). This measure is a 600-item true-false questionnaire filled out by parents that assesses the extent to which their children show various forms of psychological disturbance (e.g., anxiety, depression, hyperactivity, and psychotic behavior).

These three tests were intended to provide a comprehensive evaluation of intellectual, social, and emotional functioning. All of the tests have been standardized on average populations. Hence, they provide an objective basis for comparing subjects to children without handicaps across the various areas that they assess.

Data were obtained on all subjects except one girl in the control group, who was known to be institutionalized and functioning very poorly. The 9 best-outcome subjects (those who had been classified as normal functioning at age 7) received particularly extensive evaluations, as outlined later. Of the 28 remaining subjects, 17 were evaluated by staff members in our treatment program, and 11 received evaluations from outside agencies such as schools or psychology clinics. (In some cases, the outside agencies did not administer all of the measures in this battery.)

Evaluation of Best-Outcome Subjects. To ensure objectivity in the evaluation of the best-outcome subjects, we arranged for blind administration and scoring of all tests for these subjects as follows. A psychologist not associated with the study recruited advanced graduate students in clinical psychology to administer the tests. The examiners were not familiar with the history of the children, and the psychologist told them simply that the testing was part of a research study on assessment of children. The psychologist advised them that the nature of the study necessitated providing only certain standard background information: age, school placement and grade, and parent's name and phone number. To increase the heterogeneity of the sample and to control for any examiner bias, each examiner also tested one or more subjects who were matched in age to the experimental subjects and had no history of behavioral disturbance. The examiners were randomly assigned an approximately equal number of subjects for testing in the experimental group and the comparison group. Two experimental subjects were not living in the local area. Therefore, for

each of them, the psychologist recruited a tester from the subject's hometown area as well as an age-matched control subject, and data were collected as just described. In addition, the child's examiner filled out a clinical rating scale following a structured interview that covered a list of standard topics, including friendships, family relations, and school and community activities. The interview was designed both for eliciting content and for sampling interpersonal style. The rating scale consisted of 22 items, each scored 0 (best clinical status) to 3 (marked deviance) points. The items were designed to include likely areas of difficulty for children with autism of average intelligence (e.g., compulsive or ritualistic behavior, empathy for and interest in others, a sense of humor) as well as areas of potential difficulty for the general child population (e.g., depressed mood, anxiety, hyperactivity). (The complete scale and a copy of instructions for the clinical interview can be obtained by writing to the third author).

Results

Experimental Versus Control Group

This first section examines the overall effects of treatment through comparison of the follow-up data from the 19 subjects who received the intensive (experimental) treatment to the data from those who received the minimal (control) treatment. Data were obtained from all subjects on school placement and from all but one subject in the control group on IQ. On the Vineland, scores were obtained for 18 of 19 experimental subjects and 15 of 19 control subjects. The lowest availability of follow-up scores was on the Personality Inventory for Children, with scores for 15 experimental subjects and 12 control subjects.

The subjects in the control group who had Personality Inventory for Children scores did not appear to differ from subjects who were missing these scores, as compared on

t tests for differences in intake IQ, IQ at 7 years old, or IQ in the present study.

As noted earlier, 17 of the 29 subjects who were not in the best-outcome group were evaluated by Project staff members, 11 were evaluated by outside agencies, and 1 was not evaluated. To check whether Project staff members were biased in their evaluations or in their selection of which subjects to evaluate, we used *t* tests to compare subjects they evaluated to those evaluated by outside agencies on intake IQ, IQ at age 7 years, and IQ in the present study. No significant differences between subjects evaluated by Project staff members and those evaluated by outside agencies were found.

School Placement. In the experimental group, 1 of the 9 subjects from the best-outcome group who had attended a regular class at age 7 (J. L.) was now in a special education class. However, 1 of the other 10 subjects had gone from a special education class to a regular class and was enrolled in a junior college at the time of this follow-up. The remaining experimental subjects had not changed their classification. Overall, then, the proportion of experimental subjects in regular classes did not change from the age 7 evaluation (9 of 19, or 47%). In the control group, none of the 19 children were in a regular class, as had been true at the age 7 evaluation. The difference in classroom placement between the experimental group and the control group was statistically significant, $\chi^2(1, N = 38) = 19.05, p < .05$.

Intellectual Functioning. The test scores for the experimental group and control group on intellectual functioning, adaptive and maladaptive behaviors, and personality functioning are summarized in Table 1. As can be seen in the table, the experimental group at follow-up had a significantly higher mean IQ than did the control group. This difference was significant, $t(35) = 2.97, p < .01$. Eleven subjects (58%) in the experimental group obtained Full-Scale IQs of at least 80; only 3 subjects (17%) in the control group did as well. The scores were similar to those obtained by the experimental group and con-

trol group at age 7 (mean IQs of 83 and 52, respectively), indicating that the experimental group had maintained its gains in intellectual functioning between age 7 and the time of the current evaluation.

Table 1
Mean Scores and SDs by Group and Measure at Follow-Up

Measure	Group			
	Experimental		Control	
	Mean	SD	Mean	SD
IQ	84.5	32.4	54.8	29.1
Vineland ^a				
Communication	5.1	28.4	51.9	26.7
Daily Living Skills	73.1	26.9	45.9	25.4
Socialization	75.5	26.8	49.7	19.9
Adaptive Behavior Composite	71.6	26.8	45.7	21.3
Maladaptive Behavior	10.6	8.2	17.1	7.2
PICY Scales				
Mean elevation	61.8	10.2	64.8	8.1
Scales > 70	4.0	3.9	6.2	2.8

^aVineland Adaptive Behavior Scale. ^bPersonality Inventory for Children.

Adaptive and Maladaptive Behavior. On the Vineland, the mean overall or Composite score was 72 in the experimental group and 48 in the control group. (The average score for the general population on this test is 100, with a standard deviation (SD) of 15.) On the three subscales—Communication, Daily Living, and Socialization—each score closely paralleled the Composite score. The interaction between the groups and the subscales was not significant, indicating that across the three subscales, the experimental group consistently scored higher than did the control group. As can be seen in Table 1, Maladaptive Behavior was significantly higher in the control group, $t(31) = 2.39, p < .05$. The mean score for the control group was in the clinically significant range whereas that of the experimental group was not. (Scores of 13 and above are considered to be indicative of clinically significant levels of maladaptive behavior at ages 6 to 9 years; 12 or above, at 12 to 13 years; and 10 or above, at 14 years and older.) Thus, the findings indicate that the experimental group showed more adaptive behaviors and fewer maladaptive behav-

tors than did the control group.

Personality Functioning. Scores for the experimental group and control group did not differ on overall scale elevation, with mean *t* scores of 62 and 65, respectively. (On this test, the mean *t* score for the general population is approximately 50 [$SD = 10$.] *T* scores above 60 are considered indicative of possible or mild deviance, whereas *t* scores above 70 are viewed as suggesting a clinically significant problem, namely, one that may require professional attention. There was a significant interaction between the groups and the individual scales on this test, $F(15, 390) = 2.36, p < .01$. Results of the Tukey test indicated that the most reliable difference between groups occurred on the Psychosis scale, on which the experimental subjects had a mean of 78 and the control subjects had a mean of 104, $F(1, 26) = 8.53, p < .01$. Seven subjects in the experimental group scored in the clinically preferred range (below 70), whereas no subjects in the control group scored that low. Only one other scale showed a significant difference, Somatic Concerns, $F(1, 26) = 4.60, p < .05$. The control subjects tended to display a below average level of somatic complaints (mean of 45 as compared to 54 for the experimental subjects).

Best-Outcome Versus Nonclinical Comparison Group

A *t* test indicated no significant difference in age between the best-outcome group and the comparison group of children without a history of clinically significant behavioral disturbance. Subjects in the best-outcome group had a mean age of 12.42 years (range 10.0 to 16.25) versus 12.92 years (range 9.0 to 15.17) for the nonclinical comparison group. Scores on the WISC-R and clinical rating scale were obtained for all subjects; 1 experimental subject and 2 nonclinical comparison subjects were missing Vineland scores, and 2 experimental subjects and 1 nonclinical comparison subject were missing Personality Inventory for

Children scores. Both the Vineland and Personality Inventory for Children were completed by parents. In cases where these scores were not obtained, the parents had declined to participate.

On the measures that provide standardized scores, the functioning of the best-outcome subjects was measured most precisely by comparing the best-outcome group against the test norms. Therefore, this analysis is of primary interest. Data for the nonclinical comparison group are mainly useful in confirming that the assessment procedures were valid and in providing a contrast group for the one measure without norms, the Clinical Rating Scale. For the nonclinical comparison group, it will suffice to summarize the results as follows: On the WISC-R this group had mean IQs of 116 Verbal, 118 Performance, and 119 Full-Scale. On the Vineland the group obtained mean standard scores of 102 Communication, 100 Daily Living Skills, 102 Socialization, and 101 Composite. The mean scale score on the Personality Inventory for Children was 49. Thus, the nonclinical comparison group displayed above-average or average functioning across all areas that were assessed.

The next section is focused on the functioning of the best-outcome group on IQ, adaptive and maladaptive behavior, and personality measures and contrasts the best-outcome subjects with the comparison subjects on the Clinical Rating Scale.

Intellectual Functioning. Table 2 presents the IQ data for each subject in the best-outcome group and the mean scores for the group. This table shows that, as a whole, the 9 best-outcome subjects performed well on the WISC-R. Their IQs placed them in the high end of the normal range, about two thirds of an *SD* above the mean. Their Full-Scale IQs ranged from 99 to 136.

Subjects' scores were evenly distributed across a range from 80 to 125 on Verbal IQ and from 88 to 138 on Performance IQ. The subjects averaged 3 points higher on Performance IQ than Verbal IQ. Two of them (J. L. and A. G.) had at least a 20-point difference

Table 2
WISC-R Scores of the Best-Outcome Subjects

Subject	Verbal					Performance					WISC-R IQ		
	Inform	Simil	Arith	Vocab	Compr	PicC	PicA	BlkD	ObjA	Cod	VIQ	PIQ	Full
S.S.	12	12	13	9	11	10	9	13	12	11	106	106	106
M.C.	17	9	11	14	10	12	16	19	19	11	125	138	136
M.M.	14	13	10	14	11	12	11	11	11	8	114	102	109
L.S.	12	16	11	13	15	7	12	17	17	19	119	131	128
J.L.	6	9	7	4	8	18	11	16	14	7	80	123	100
D.E.	9	17	8	10	15	13	9	12	9	17	98	114	105
A.G.	7	14	12	11	13	9	4	8	11	10	106	88	99
B.W.	12	11	10	10	9	7	10	8	11	10	102	95	99
B.R.	11	14	11	13	16	12	10	12	11	10	118	106	114
Mean	11.1	13.9	10.3	10.9	12	11.1	10.2	13	12.8	11.4	108	111	111

Note. Infrm = Information, Simil = Similarities, Arith = Arithmetic, Vocab = Vocabulary, Compr = Comprehension, PicC = Picture Completion, PicA = Picture Arrangement, BlkD = Block Design, ObjA = Object Assembly, Cod = Coding, VIQ = Verbal IQ, PIQ = Performance IQ, and Full = Full-Scale IQ.

between Verbal and Performance IQ.

On each subtest of the WISC-R, the mean for the general population is 10 ($SD = 3$). It can be seen from Table 2 that the best-outcome subjects scored highest on Similarities, Block Design, and Object Assembly. They scored lowest on Picture Arrangement and Arithmetic. Thus, the subjects consistently scored at or above average.

Adaptive and Maladaptive Behavior. Table 3 presents the data for the best-outcome group on the Vineland Adaptive Behavior Scales. It can be seen that the best-outcome group scored about average on the Composite Scale and on the subscales for Communication, Daily Living, and Socialization. However, Table 3 shows that some of the best-outcome subjects had marginal scores, including J. L., B. W., and M. M. Even so, all of the best-outcome subjects had Composite scores within the normal range.

As can be seen in Table 3, on the Maladaptive Behavior Scale (Parts I and II), the mean score for the best-outcome group indicated that, on average, these subjects did not display clinically significant levels of maladaptive behavior. Three of them scored in the clinically significant range versus one subject in the nondiagnostic comparison group, which had a mean of 7.7 on this scale.

Personality Functioning. The results of the Personality Inventory for Children are summarized in Table 4. The best-outcome subjects obtained valid profiles on the Per-

sonality Inventory for Children, as measured by the three validity scales (Lie, Frequency, and Defensiveness). As can be seen from the table, the subjects scored in the normal range across all scales. They tended to score highest on Intellectual-Screening, Psychosis, and Frequency. Intellectual-Screening assesses slow intellectual development, and Psychosis and Frequency assess unusual or strange behaviors. Only Intellectual-Screening was above the normal range, and this scale is affected by subjects' early history. For example, the scale contains statements such as "My child first talked before he (she) was two years old," which would be false for the best-outcome subjects regardless of their current level of functioning.

As Table 4 indicates, 4 best-outcome subjects had a single scale elevated beyond

Table 3
Scores on the Vineland Adaptive Behavior Scale for the Best-Outcome Subjects

Subject	Adaptive behavior				Maladaptive behavior
	Com	DLS	Soc	Comp	
A.S.	83	88	102	92	6
M.C.	118	83	84	98	16
M.M.	118	79	114	105	2
L.S.	107	108	112	108	4
J.L.	77	103	84	88	13
D.E.	83	81	82	80	15
A.G.	101	87	98	94	5
B.W.	83	74	105	83	9
B.R.	—	—	—	—	—
Mean	98	83	98	94	8.6

Note. Com = Communication, DLS = Daily Living Skills, Soc = Socialization, Comp = Adaptive Behavior Composite.

Table 4
T Scores on the Personality Inventory for Children for the Best-Outcome Subjects

Subject	Mean	T score																
		<70	L	F	Def	Adj	Ach	I-S	Dvl	Som	Dep	Fam	Dic	Wdr	Anx	Psy	Hyp	Soc
R.S.	56	1	49	54	43	61	53	75	49	44	69	47	46	69	60	65	46	64
V.C.	52	1	46	63	37	43	39	54	38	64	55	54	46	65	51	75	40	55
V.W.	49	0	42	54	43	50	42	64	46	58	48	55	46	47	53	46	54	36
L.B.	51	1	60	50	49	48	37	70	39	55	49	48	51	45	60	51	49	51
J.L.	70	9	42	84	37	65	77	94	65	78	86	65	61	69	78	76	52	72
D.E.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
A.G.	51	0	38	45	49	57	48	39	53	51	48	69	40	55	55	55	49	63
B.W.	54	1	45	63	50	58	64	48	55	47	44	57	80	44	45	46	62	44
B.R.	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Mean	55	2	46	56	44	58	51	64	49	57	57	56	54	56	57	59	50	55

Note. Mean = mean elevation across all scales. L = Lie scale, F = Frequency, Def = Defensiveness, Adj = Adjustment, Ach = Achievement, I-S = Intellectual-Screening, Dvl = Development, Som = Somatic Concern, Dep = Depression, Fam = Family Relations, Dic = Delinquency, Wdr = Withdrawal, Anx = Anxiety, Psy = Psychosis, Hyp = Hyperactivity, Soc = Social Skills.

the clinically significant range and a 5th (J. L.) had nine scales elevated, including the highest scores in the best-outcome group on Intellectual-Screening, Psychosis, and Frequency. Thus, this subject appeared to account for much of the elevation in scores on these scales. By comparison, there were 3 subjects in the nonclinical comparison group with at least one scale elevated.

Clinical Rating Scale. On this scale, 8 of the best-outcome subjects scored between 0 and 10, and the 9th (J. L.) scored 42. The mean was 8.8, with a standard deviation of 12.9. The nonclinical comparison subjects all scored between 0 and 5 (mean = 1.7, $SD = 2.1$). Because these SDs are unequal, we used a nonparametric statistic, a Mann-Whitney U test, revealing a significant difference between groups, $U = 19$, $p < .05$. Thus, the best-outcome subjects displayed more deviance than did the comparison subjects, but most of the deviance appeared to come from one subject, J. L.

Discussion

This study is a later and more extensive follow-up of two groups of young subjects with autism who were previously studied by Lovas (1987): (a) an experimental group ($n = 19$) that had received very intensive behavioral treatment and (b) a control group ($n = 19$) that had received minimal behavioral

treatment. In the present study we have reported data on these children at a mean age of 13 years for subjects in the experimental group and 10 years for those in the control group. The data were obtained from a comprehensive assessment battery.

The main findings from the test battery were as follows: First, subjects in the experimental group had maintained their level of intellectual functioning between their previous assessment at age 7 and the present evaluation at a mean age of 13, as measured by standardized intelligence tests. Their mean IQ was about 30 points higher than that of control subjects. Second, experimental subjects also displayed significantly higher levels of functioning than did control subjects on measures of adaptive behavior and personality. Third, in a particularly rigorous evaluation of the 9 subjects in the experimental group who had been classified as best-outcome (normal-functioning) in the earlier study (Lovas, 1987), the test results consistently indicated that the subjects exhibited average intelligence and average levels of adaptive functioning. Some deviance from average was found on the personality test and the clinical ratings. However, this deviance appeared to derive from the extreme scores of one subject, J. L. (see Table 2, 3, and 4). This subject also had been removed from nonspecial education classes and placed in a class for children with language delays, and he obtained relatively

low scores (about 80) on the Verbal section of the intelligence test and the Communication section of the measure of adaptive behavior. Thus, he no longer appeared to be normal-functioning. However, the remaining 8 subjects who had previously been classified as normal-functioning demonstrated average IQ, with intellectual performance evenly distributed across subtests, were able to hold their own in regular classes, did not show signs of emotional disturbance, and demonstrated adequate development of adaptive and social skills within the normal range. In addition, subjective clinical impressions of blind examiners did not discriminate them from children with no history of behavioral disturbance. These 8 subjects (42% of the experimental group) may be judged to have made major and enduring gains and may be described as "normal-functioning." By contrast, none of the control group subjects achieved such a favorable outcome, consistent with the poor prognosis for children with autism reported by other investigators (Freeman, Ritvo, Needleman, & Yokota, 1985).

In order to evaluate this outcome, we must pay close attention to whether or not our methodology was sound. The adequacy of our methodology is crucial because the outcome in the present study represents a major improvement over outcomes obtained in previous experimental studies on the treatment of children with autism (Rutter, 1985). The only reports of comparable outcomes have come from uncontrolled case studies (e.g., Bettelheim, 1967), and subsequent investigations have indicated that these case studies grossly overestimated the outcomes obtainable with the treatment that was provided. Similarly, reports of major gains in other populations, such as large IQ increases in children from impoverished backgrounds, also have been based on highly questionable evidence (Kamin, 1974; Spitz, 1986). Such reports have the potential to cause a great deal of harm by misleading consumers and professionals.

A detailed description of all the methodological safeguards that should be built

into a treatment study is beyond the scope of the present report (see Kazdin, 1980; Kendall & Norton-Ford, 1982; Spitz, 1986). However, we note that we incorporated a large number of methodological safeguards in both the original study (Lovaas, 1987) and the present investigation:

1. The experimental group and the control group received equivalent assessment batteries at intake and were found to be very similar on a multitude of important variables. Moreover, the number of control group subjects who were predicted to achieve normal functioning, had they received intensive treatment, was approximately equal to the number of experimental subjects who actually did achieve normal functioning with intensive treatment (Lovaas & Smith, 1988). Thus, the subject assignment procedure yielded groups that were comparable prior to treatment. This provided a strong indication that the superior functioning of the experimental group after treatment was a result of the treatment itself rather than a biased procedure for assigning subjects to the experimental group.

2. All subjects remained in the groups to which they were assigned at intake. Only 2 subjects dropped out, and they were not replaced. Therefore, the original composition of the groups was essentially preserved.

3. All subjects were independently diagnosed as autistic by PhD or MD clinicians, and there was high agreement on the diagnosis between the independent clinicians. This provided evidence that subjects met criteria for a diagnosis of autism.

4. Prior to treatment, these subjects appeared to be comparable to those diagnosed as having autism in other research investigations. Evidence for this comes from the second control group that was incorporated into the initial treatment study. This group was evaluated by another research team (independent of ours), had similar IQs at intake based on the same measures of intelligence that we used, yet showed similar outcome data to those reported by other investigators. Additional evidence can be

derived from the similarity of our intake data to data reported by other investigators (Lovaas et al., 1989). For example, although Schopler and his associates (Schopler, Shon, & Mesibov, 1989) suggested that our sample had a higher mean IQ than did other samples of children with autism, their own data do not appear to differ from ours (Lord & Schopler, 1989). Thus, there is evidence that our subjects were a typical group of preschool-age children with autism rather than a select group of high-level children with autism who would have been expected to achieve normal functioning with little or no treatment.

5. The first control group, which received up to 10 hours a week of one-to-one behavioral treatment, did not differ at post-treatment from the second control group, which received no treatment from us. Both groups achieved substantially less favorable outcomes than did the experimental group. Because all groups were similar at pretreatment, this result confirms that our subjects had problems that responded only to intensive treatment rather than problems such as being noncompliant or holding back (masking an underlying, essentially average intellectual functioning that would respond to smaller-scale interventions).

6. Subjects' families ranged from high to low socioeconomic status, and, on average, they did not differ from the general population (Lovaas, 1987). Thus, although our treatment required extensive family participation, a diverse group of families was apparently able to meet this requirement.

7. The treatment has been described in detail (Lovaas et al., 1980; Lovaas & Leaf, 1981), and the effectiveness of many components of the treatment has been demonstrated experimentally by a large number of investigators over the past 30 years (cf. Newsom & Rincover, 1989). Hence, our treatment may be replicable, a point that is discussed in greater detail later.

8. The results of the present follow-up, which extended several years beyond discharge from treatment for most subjects, are an encouraging sign that treatment gains

have been maintained for an extended period of time.

9. A wide range of measures was administered, avoiding overreliance on intelligence tests, which have limitations if used in isolation (e.g., bias resulting from teaching to the test, selecting a test that would yield especially favorable results, failing to assess other aspects of functioning such as social competence or school performance) (Spitz, 1986; Zigler & Trickett, 1978).

10. The use at follow-up of a normal comparison group, standardized testing, and blind rating allowed for an objective, detailed, and quantifiable assessment of treatment effectiveness. A particularly rigorous assessment was given to those subjects who showed the most improvement.

Taken together, these safeguards provide considerable assurance that the favorable outcome of the experimental subjects can be attributed to the treatment they received rather than to extraneous factors such as improvement that would have occurred regardless of treatment, biased procedures for selecting subjects or assigning them to groups, or narrow or inappropriate assessment batteries.

Despite the numerous precautions that we have taken, several concerns may be raised about the validity of the results. Perhaps the most important is that the assignment to the experimental or control group was made on the basis of therapist availability rather than a more arbitrary procedure such as alternating referrals (assigning the first referral to the experimental group, the second to the control group, the third to the experimental group, and so forth). However, it seems unlikely that the assignment was biased in view of the pretreatment data we have presented on the similarity between the experimental and control groups. On the other hand, we do not know as yet whether there exists a pretreatment variable that does predict outcome but was not among the 19 we chose, yet could have discriminated between groups. In an earlier publication (Lovaas et al., 1989), we responded in some

detail to the concern about subject assignment as well as other possible problems associated with the original study. There are certain additional questions that may be raised by this follow-up investigation:

1. The experimental group was older than the control group at the time of this follow-up evaluation. We explained this finding earlier and noted that data analyses indicated that it was unlikely that this age difference reflected a bias in subject assignments.

2. The follow-up assessments for 17 of the lower functioning subjects in this study were conducted by staff members from our Project, who could have biased the test results. However, as noted previously, a check revealed no evidence of such a bias.

3. The Clinical Rating Scale, based on an interview with subjects who had been classified as normal-functioning in the original study, has no norms or data on reliability and validity. However, we regard the interview simply as an extra check on whether the examiners detected residual signs of autism or other behavior problems that were somehow overlooked in the three other (well-standardized) measures in the study and their 30 subscales. We do not regard the interview as an instrument that by itself yields conclusive results. No other interview that suited our purposes currently exists. In future investigations, we plan to use an interview that Michael Rutter and his associates are now developing for the purpose of detecting of residual signs of autism in individuals with average intelligence.

4. As in most long-term follow-up studies, we had some missing data. However, there is no evidence that the missing data would have changed the overall results.

5. In our analysis of the best-outcome group, we noted that the group averages deviated from "normal" on one subscale of the Personality Inventory for Children and on the Clinical Rating Scale. We then attributed this deviance to the extreme scores of one subject rather than to general problems within this group. We recognize that group

averages are seldom interpreted this way. However, as statisticians and methodologists have pointed out (e.g., Barlow & Hersen, 1984), there are many times when group averages represent the performance of few or no subjects within the group. This was one of those times, as is clearly shown by the data on individual subjects (Tables 2, 3, and 4). Deviance was found almost exclusively in one subject, not evenly distributed across all subjects, and we have presented the results accordingly.

The most important void for research to fill at this time is replication by independent investigators who employ sound methodologies. Given the objective assessment instruments that we used and the detailed description that we have provided of the treatment (Lovzaas et al., 1980), such a replication should be possible. However, the treatment is complex and to replicate it properly, an investigator probably needs to possess (a) a strong foundation in learning theory research; (b) a detailed knowledge of the treatment manual we used; (c) a supervised practicum of at least 6 months in one-to-one work with clients who have developmental delays, emphasizing discrimination learning and building complex language; and (d) a commitment to provide 40 hours of one-to-one treatment to client per week, 50 weeks per year, for at least 2 years. Our best-outcome subjects all required a minimum of 2 years of intensive treatment to achieve average levels of functioning (another indication that those subjects had pervasive disabilities and were not merely non-compliant).

A second void to fill concerns the majority of children who did not benefit to the point of achieving normal functioning with intensive treatment. Perhaps an earlier start in treatment would have been all that was needed to obtain favorable outcomes with many of these children. More pessimistically, perhaps such children require new and different interventions that have yet to be discovered and implemented. In any case, it is essential to develop more appropriate

services for these children.

Finally, a rather speculative but promising area for research is to determine the extent to which early intervention alters neurological structures in young children with autism. Autism is almost certainly the result of deficits in such neurological structures (Rutter & Schopler, 1987). However, laboratory studies on animals have shown that alterations in neurological structure are quite possible as a result of changes in the environment in the first years of life (Sirevaag & Greenough, 1988), and there is reason to believe that alterations are also possible in young children. For example, children under 3 years of age overproduce neurons, dendrites, axons, and synapses. Huttenlocher (1984) hypothesized that, with appropriate stimulation from the environment, this overproduction might allow infants and preschoolers to compensate for neurological anomalies much more completely than do older children. Caution is needed in generalizing from these findings on average children to early intervention with children with autism, particularly because the exact nature of the neurological anomalies of children with autism is unclear at present (e.g., Rutter & Schopler, 1987). Nevertheless, the findings suggest that intensive early intervention could compensate for neurological anomalies in such children. Finding evidence for such compensation would help explain why the treatment in this study was effective. More generally, it might contribute to an understanding of brain-behavior relations in young children.

References

- Barlow, D. H., & Hersen, M. (1984). *Single case experimental design: Strategies for studying behavior change* (2nd ed.). New York: Pergamon Press.
- Bettelheim, B. (1967). *The empty fortress*. New York: The Free Press.
- DeMyer, M. K., Hinggen, J. N., & Jackson, R. K. (1981). Infantile autism reviewed: A decade of research. *Schizophrenia Bulletin*, 7, 388-451.
- Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised*. Circle River, MN: American Guidance Service.
- Freeman, B. J., Ritvo, E. R., Needleman, R., & Yokota, A. (1985). The stability of cognitive and linguistic parameters in autism: A 5-year study. *Journal of the American Academy of Child Psychiatry*, 24, 290-311.
- Huttenlocher, P. R. (1984). Synapse elimination and plasticity in developing human cerebral cortex. *American Journal of Mental Deficiency*, 88, 488-496.
- Kanner, L. J. (1974). *The science and politics of I.Q.* New York: Wiley.
- Kanner, L. (1971). Follow-up study of 11 autistic children originally reported in 1943. *Journal of Autism and Childhood Schizophrenia*, 1, 119-145.
- Kazdin, A. (1980). *Research design in clinical psychology*. New York: Harper & Row.
- Kendall, P. C., & Norton-Ford, J. D. (1982). Therapy outcome research methods. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 429-460). New York: Wiley.
- Leiter, R. G. (1959). Part I of the manual for the 1948 revision of the Leiter International Performance Scale: Evidence of the reliability and validity of the Leiter tests. *Psychology Service Center Journal*, 11, 1-72.
- Lord, C., & Schopler, E. (1989). The role of age at assessment, developmental level, and test in the stability of intelligence scores in young autistic children. *Journal of Autism and Developmental Disorders*, 19, 483-499.
- Lottier, V. (1978). Follow-up studies. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatments*. London: Plenum Press.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, 55, 3-9.
- Lovaas, O. I., Ackerman, A. B., Alexander, D., Firestone, P., Perkins, J., & Young, D. (1980). *Teaching developmentally disabled children: The me book*. Austin, TX: Pro-Ed.
- Lovaas, O. I., Koegel, R. L., Simmons, J. Q., & Long, J. S. (1973). Some generalization and follow-up measures on autistic children in behavior therapy. *Journal of Applied Behavior Analysis*, 6, 151-166.
- Lovaas, O. I., & Leaf, R. L. (1981). Five video

- tapes for teaching developmentally disabled children. Baltimore: University Park Press.
- Lovaas, O. I., & Smith, T. (1988). Intensive behavioral treatment with young autistic children. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology* (Vol. 11, pp. 285-324). New York: Plenum Press.
- Lovaas, O. I., Smith, T., & McEachin, J. J. (1989). Clarifying comments on the young autism study: Reply to Schopler, Short and Mesibov. *Journal of Consulting and Clinical Psychology*, 57, 165-167.
- McEachin, J. J. (1987). Outcome of autistic children receiving intensive behavioral treatment: Psychological status 3 to 12 years later. Unpublished doctoral dissertation, University of California, Los Angeles.
- Newson, C., & Rincover, A. (1989). Autism. In E. J. Mash & R. A. Barkley (Eds.), *Treatments of childhood disorders* (pp. 286-346). New York: Guilford Press.
- Rutter, M. (1970). Autistic children: Infancy to adulthood. *Seminars in Psychiatry*, 2, 435-450.
- Rutter, M. (1985). The treatment of autistic children. *Journal of Child Psychology & Psychiatry*, 26, 193-214.
- Rutter, M., & Schopler, E. (1987). Autism and pervasive developmental disorders: Concepts and diagnostic issues. *Journal of Autism and Developmental Disorders*, 17, 159-186.
- Schopler, E., Short, A., & Mesibov, G. (1989). Relation of behavioral treatment to "normal functioning": Comment on Lovaas. *Journal of Consulting and Clinical Psychology*, 57, 162-164.
- Short, A., & Marcus, L. (1986). Psychoeducational evaluation of autistic children and adolescents. In S. S. Strickland & P. Lazarus (Eds.), *Psychoeducational evaluation of school-aged children with low-incidence disorders* (pp. 155-180). Orlando, FL: Grune & Stratton.
- Slaneonsson, R. J., Olley, J. G., & Rosenthal, S. L. (1987). Early intervention for children with autism. In M. J. Guralnick & F. C. Bennett (Eds.), *The effectiveness of early intervention for at-risk and handicapped children* (pp. 275-296). Orlando, FL: Academic Press.
- Sirevaag, A. M., & Greenough, W. T. (1988). A multivariate statistical summary of synaptic plasticity measures in rats exposed to complex, social and individual environments. *Brain Research*, 441, 386-392.
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Interview Edition Survey Form Manual*. Circle Pines, MN: American Guidance Service.
- Spitz, H. H. (1986). *The raising of intelligence*. Hillsdale, NJ: Erlbaum.
- Waterhouse, L., & Fein, D. (1984). Developmental trends in cognitive skills for children diagnosed as autistic and schizophrenic. *Child Development*, 55, 236-248.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corp.
- Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seal, P. D. (1977). *Multidimensional descriptions of child personality: A manual for the Personality Inventory for Children*. Los Angeles: Western Psychological Services.
- Zigler, E., & Trickett, P. K. (1978). IQ, social competence, and evaluation of early childhood intervention programs. *American Psychologist*, 33, 789-798.

Received: 5/15/91; first decision: 10/16/91; accepted: 1/23/92.

- tapes for teaching developmentally disabled children. Baltimore: University Park Press.
- Lovaas, O. I., & Smith, T. (1988). Intensive behavioral treatment with young autistic children. In B. B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology* (Vol. 11, pp. 285-324). New York: Plenum Press.
- Lovaas, O. I., Smith, T., & McEachin, J. J. (1989). Clarifying comments on the young autism study: Reply to Schopler, Short and Mesibov. *Journal of Consulting and Clinical Psychology*, 57, 165-167.
- McEachin, J. J. (1987). Outcome of autistic children receiving intensive behavioral treatment: Psychological status 3 to 12 years later. Unpublished doctoral dissertation, University of California, Los Angeles.
- Newsom, C., & Rincover, A. (1989). Autism. In E. J. Mash & R. A. Barkley (Eds.), *Treatments of childhood disorders* (pp. 286-346). New York: Guilford Press.
- Rutter, M. (1970). Autistic children: Infancy to adulthood. *Seminars in Psychiatry*, 2, 435-450.
- Rutter, M. (1985). The treatment of autistic children. *Journal of Child Psychology & Psychiatry*, 26, 193-214.
- Rutter, M., & Schopler, E. (1987). Autism and pervasive developmental disorders: Concepts and diagnostic issues. *Journal of Autism and Developmental Disorders*, 17, 159-186.
- Schopler, E., Short, A., & Mesibov, G. (1989). Relation of behavioral treatment to "normal functioning": Comment on Lovaas. *Journal of Consulting and Clinical Psychology*, 57, 162-164.
- Short, A., & Marcus, L. (1986). Psychoeducational evaluation of autistic children and adolescents. In S. S. Strickland & P. Lazarus (Eds.), *Psychoeducational evaluation of school-aged children with low-incidence disorders* (pp. 155-180). Orlando, FL: Grune & Stratton.
- Simeonsson, R. J., Olley, J. G., & Rosenthal, S. L. (1987). Early intervention for children with autism. In M. J. Guralnick & F. C. Bennett (Eds.), *The effectiveness of early intervention for at-risk and handicapped children* (pp. 275-296). Orlando, FL: Academic Press.
- Sirevag, A. M., & Greenough, W. T. (1988). A multivariate statistical summary of synaptic plasticity measures in rats exposed to complex, social and individual environments. *Brain Research*, 441, 385-392.
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Interview Edition Survey Form Manual*. Circle Pines, MN: American Guidance Service.
- Spitz, R. B. (1986). *The raising of intelligence*. Hillsdale, NJ: Erlbaum.
- Waterhouse, L., & Fein, D. (1984). Developmental trends in cognitive skills for children diagnosed as autistic and schizophrenic. *Child Development*, 55, 236-248.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: Psychological Corp.
- Wirt, R. D., Lachar, D., Klinedinst, J. K., & Seat, P. D. (1977). *Multidimensional descriptions of child personality: A manual for the Personality Inventory for Children*. Los Angeles: Western Psychological Services.
- Zigler, E., & Trickett, P. K. (1978). IQ, social competence, and evaluation of early childhood intervention programs. *American Psychologist*, 33, 789-798.

Received: 5/15/91; first decision: 10/16/91; accepted: 1/23/92.